FFCUL/LASIGE FINAL REPORT

The Epiwork project started in February 2009 and run for 54 months (48 months + 6 months extension). According to the work plan FFCUL is involved in the following work packages:

- WP 1 Population Models and Contact Networks
- WP 3 Information platform
- WP 4 Epidemic Modelling Platform
- WP 7 Management

The FFCUL participation in Epiwork involves two groups:

- CMAF Group
- LASIGE Group.

This report summarises the progress in the three last semesters of the project activity by the LASIGE Group, which leads WP3 and participates in WP4 and WP7.

In the reporting period (forth year plus extension semester, ie from February 1, 2012 to July 31, 2013), the commitment was 86.99 persons.month, 77.55 pm from technicians, and 9.45 pm from permanent staff.

The total commitment of the FFCUL LASIGE Group to Epiwork was 223.69 persons.month, 188.43 from technicians and 35.27 from permanent staff.

The LASIGE Team working on WP3 and WP4 in the reporting period includes:

- Mário J. Silva (IST Faculty, worked in the reporting period in WP3, WP4 and WP7)
- Francisco Couto (FCUL Faculty, worked in the reporting period in WP3)
- Dulce Domingos (FCUL Faculty, worked in the reporting period in WP3 and WP4)
- Catia Pesquita (FCUL Non-Permanent Faculty worked half-time for 3 months in WP3)
- Paulo Graça (software developer, working full-time in WP3 and WP4 until June 30, 2013)
- João Zamite (Graduate student, worked full-time in the reporting period).
- João Ferreira (Graduate student, worked full-time for nine months during the forth year).
- Carlos Sousa (Graduate student, worked full-time in the reporting period)
- Tiago Posse (Graduate student, worked full-time in the reporting period)
- Vera Carvalho (Undergraduate student, worked in the Summer 13)

The following changes occurred in the FFCUL/LASIGE Team:

Cátia Pesquita, who worked as a graduate research student in the research team, on the production of the Deliverable D3.5, worked half-time for part of the forth year on the evaluation of NERO ontologies for Epiwork datasets annotation after completing her PhD (the remainder as Adjunct Professor at the University).

João Ferreira worked full-time on implementing the annotation of datasets with NERO concepts and its implementation in Epiwork's user interface.

Vera Carvalho was hired with a scholarship as a summer intern for the manual annotation of Epiwork datasets.

Paulo Graça left the project at the end of M53 (June 2013).

1. WP3 — Information platform: Work progress, collaborations and achievements during the period

This Work Package is lead by FFCUL, with a total contribution of 82 persons.month (60 hired technicians + (22) academics).

2.1 *Background Information on WP3*

Work package number	3		**Start date or starting event:**		Month 1		
Work package title	**Information platform**						
Activity type	RTD						
Participant number	12	1	2	3	4	5	9	10
Participant short name	FFCUL	ISI	FGC-IGC	TAU	MPI-DS	AIBV	BIU	FBK-IRST
Person-months per participant	167	60	4	8	15	4	11	6

The person-monts per participant shown above correspond to the new estimate after included in the amendment to the Description of Work.

This Work Package was lead by FFCUL. The whole WP3 activity was structured into four tasks:

Task 3.1 – Data Collection.

Participants: FFCUL, ISI, FBK-IRST, BIU, MPI-DS, FGC-IGC, AIBV.

Description: Realistic simulations of epidemic processes crucially depend on the availability of datasets describing human behaviour and pathogen-host interactions. Datasets include population movement data, social and behavioural data, health related data, geographic data, detailed geo-

temporal epidemic incidence and immunization data, pathogen evolution and multi-strains circulation data. Data can come from a variety of different sources, including hospital records, country statistics, Web content, and others. It can range from a global scale, such as the worldwide air transportation infrastructure, down to the detailed description of individual activities at a minute-by-minute scale. This task will create a catalogue of databases of epidemiological data across Europe, with extensive meta-data describing the main characteristics of the available information sources. This catalogue will be integrated with a collaborative platform that will be set up for online discussion and exchange of meta-data among the participants.

Task 3.2 – Meta-Model Design.

Participants: FFCUL, IGC, ISI.

Description: While some of the previously mentioned datasets are freely available on the Web (e.g. WHO Global Health Atlas, Eurostat), they are often scattered in different repositories, cover partial regions of the world and come in different formats, according to different standards and classifications. The project envisions a unified and integrated approach for the management of these resources, with the design and implementation of an Epidemic Marketplace Platform, publicly available on the web. The platform supports the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. It is an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on a Web2.0 approach, users will become active participants, sharing information and data, and collaborating online, rather than being satisfied with a passive information consumer/viewer role. We envision proposing a simple reference format, which will facilitate the navigation and use of the datasets. Each dataset will come with a metadata file, signalling general metadata for resource management, containing data such as: the title, the date of submission, version, the source of the data and coverage. Moreover, the metadata will include information for a more thorough description of the data included in the dataset, providing a framework for a more specific description, for example, of epidemiologic and

geographic data. The Marketplace will support flexible and intuitive tools for navigation and selection of resources. Standard classifications as well as tagging systems proposed by users will be supported.

Task 3.3 – Epidemic Marketplace Platform.

Participants: FFCUL, ISI.

Description: This task will implement a platform based on the integration of grid technology and publicly available services and software on the web to support the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. The Epidemic Marketplace Platform will be an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on some of the Web2.0 characteristics, users will become active participants, generating information and providing data for sharing, and collaborating online, rather than being satisfied with a passive information consumer/viewer role. More specifically, researchers can use and contribute to the Marketplace in several different ways. They can: (1) use it as a catalogue of data sources containing the metadata describing existing databases; (2) view, download, tag, and comment on the available resources; (3) provide compliant datasets and relevant information; (4) use it as a forum where to publish information about their own data, seek modellers to collaborate with, share and distribute their new findings.

Task 3.4 – Evaluation and monitoring of the use of the catalogue and collaboration services.

Participant: FFCUL.

Description: This task involves the monitoring of epidemiological data exchanges performed through the mediating services platform. The evaluation will assess not only the coverage of the catalogued resources, but the users' satisfaction with the user interface and integrated collaborative tools made available through the epidemiological marketplace platform. More importantly, the analysis of the collected datasets and their annotations and usage will provide a rich environment for deriving an epidemiology ontology, which will help further on the integration and communication among the community of epidemiologists.

2.2 *Progress in the Reporting Period*

- *A summary of progress towards objectives, and details for each task the team was involved;*
- *Highlight clearly significant results;*
- *Team publications within the scope of the project (please provide a pdf of the publication if possible);*
- *List of outreach activities (conferences, Invited talks, presentations, workshops, tutorials,);*
- *List of press releases or media coverage, any particular dissemination activity;*
- *unanticipated finding, opportunity etc*

Main Activities at FFCUL in the first semester of the third year of the project in WP3:

1. Mário Silva, João Zamite and Carlos Santos participated in the Epiwork Review in Brussels in March, 2012 – Mário Silva presented Deliverables D3.4 and D3..5 to the EC and gave an overview of the progress and challenges in WP3. This included the discussion of the NERO set of ontologies and reporting on the collected statistics of the usage of the EM. The demo of the first functional Epidemic Marketplace prototype with all the anticipated components was presented to the reviewers.
2. **[opportunity]** The LASIGE team maintained contacts with other teams involved in projects for the development of biomedical ontologies and terminologies. Miquel Porta, author of the popular "Dictionary of Epidemiology" is very interested in our development of an epidemiological meta-model to describe epidemiological datasets and a possible collaboration may happen. At ICBO 2012, we presented NERO and the Epidemic marketplace to Laszlo Balkanyi (ECDC), which was very interested in the project.
3. Daniela Paolotti visited LASIGE for one week in May for planning the integration of the GLEaMviz platform and the Epidemic Marketplace and discuss the integration of the EM (WP3) with Infuenzanet (WP5),

regarding the organization of datasets for use for computational epidemiology tools.

4. **[outreach]** The LASIGE team published and presented one full paper at and demonstrated the Epidemic Marketplace the International Conference of Biomedical Ontologies, in July, which was a significant step towards the disclosure of our work to this important community.
5. **[outreach]** The LASIGE team participated with three of its members in the Epiwork Digital Epidemiology workshop, with one keynote presentation, a demo of the Epidemic Marketplace and two poster communications.
6. **[opportunity]** Demos of the EM to ECDC officers at ICBO and at the Workshop on Digital Epidemiology to Pasi Penttinen (ECDC)
7. **[outreach]** Francisco Couto was invited to give a seminar at Semantic Similarity in Biomedical Ontologies: measurement, assessment and applications; 4th Workshop of Ontologies in Biomedicine and Life Sciences (OBML), Dresden, Germany, September 2012
8. Continued with the design and implementation of new access control methods for the EM. The EM access control was extended with dynamic roles.
9. Completed implementation of the final design of the EM front-end, with much improved usability, which has been online since early May 2013. Upgrade to the EM implied additional work on the integration of the Computational Platform tools, in particular GleamViz.
10. Completed integration with WP5 – anonimised data from influenza.net automatically imported to the Epidemic Marketplace and packed as epidemic datasets.

Publications and Presentations:

1. **[publication + presentation]** João D. Ferreira, Catia Pesquita, Francisco M. Couto, Mário J. Silva, Bringing epidemiology into the

Semantic Web. International Conference on Biomedical Ontologies (ICBO) 2012.

2. **[publication + presentation]** Francisco M. Couto, João D. Ferreira, João Zamite, Carlos Santos, Tiago Posse, Paulo Graça, Dulce Domingos, Mário J. Silva, The Epidemic Marketplace Platform: towards semantic characterization of epidemiological resources using biomedical ontologies. International Conference on Biomedical Ontologies (ICBO) 2012.
3. **[presentation]** João D. Ferreira, Francisco M. Couto, Semantic Similarity in the Biomedical Domain. Braga, Portugal. Poster @ Bioinformatics Open Days 2012. 2012.
4. **[publication]** Catia Pesquita, Francisco M. Couto 2012: Predicting the Extension of Biomedical Ontologies. PLOS Computational Biology 9(8), e1002630.
5. **[publication]** João Ferreira, Daniela Paolotti, Francisco Couto, Mário J. Silva 2012: On the usefulness of ontologies in epidemiologic research and practice. Journal of Epidemiology and Community Health.
6. **[publication]** Tiago Grego, Francisco Couto. Enhancement of Chemical Entity Identification in Text Using Semantic Similarity Validation. PLOS ONE 5(8), e62984. 2013.
7. **[publication]** Tiago Grego, Francisco Pinto, Francisco Couto, LASIGE: using Conditional Random Fields and ChEBI ontology. Proceedings of the International Workshop on Semantic Evaluation (SemEval). 2013.
8. **[presentation]** Catia Pesquita, João D. Ferreira, Francisco M. Couto, Mário J. Silva, Semi-Automated Annotation of Epidemiological Resources. Turin, Italy. Poster @ Epiwork International Workshop "Digital Epidemiology", May 2013.
9. **[presentation]** Carlos Santos, Dulce Domingos, João Zamite, Paulo Graça, Mário J. Silva, Access Control for Shared Epidemic Datasets. Turin, Italy. Poster @ Epiwork International Workshop "Digital Epidemiology", May 2013.

10. [publication] João Zamite, João D. Ferreira, Paulo Graça, Carlos Santos, Tiago Posse, Cátia Pesquita, Dulce Domingos, Francisco Couto, Mário J. Silva, D3.6 - Report: Final specification of the Epidemic Marketplace Platform and evaluation results Technical Report. Technical Report . University of Lisbon, Faculty of Sciences, LASIGE, July 2013.

11. [publication + presentation] João Zamite, Dulce Domingos, Mário J. Silva, Carlos Santos. Group-Based Discretionary Access Control for Epidemiological Resources. HCist'2013 - International Conference on Health and Social Care Information Systems and Technologies. To be published on Procedia Technology, Elsevier. October 2013.

12. [publication] Carlos Santos, Sistema de controlo de acesso para a Epidemic Marketplace (in Portuguese). Master Dissertation, in preparation. September 2013.

13. [publication] Catia Pesquita, João Ferreira, Vera Carvalho, Francisco M Couto, Mário J Silva. Epidemiological resource identifiers and their semantics, where are they? Submitted to PLOS One, June 2013.

14. [publication] Catia Pesquita, João D. Ferreira, Francisco M. Couto, Mário J. Silva. The Epidemiology Ontology: an ontology for the semantic annotation of epidemiology resources. Submitted to JBMS Special Isuue on Ontology, May 2013.

Activities at FFCUL in the reporting period (year 4 + extension) in Task 3.1:

- Catalogued a large set of resources described in the literature – papers from the Journal of Epidemiology
- Integration with WP5 enabled the automatic upload of datasets from Influenza.net into the Epidemic Marketplace.

Activities at FFCUL in the reporting period (year 4 + extension) in Task 3.2:

- In the period, we have extensively validated the meta-model proposed for the EM. Paper submitted to PLOS.

- Created a new Ontology "The Epidemic Ontology" (EO), which is designed to support the semantic annotation of epidemiology resources. The EO is integrated into NERO and complements epidemically relevant aspects that are not described in the existing ontologies. The EO currently contains 130 classes, 53 synonyms, 51 cross-references to seven external resources including MeSH, UMLS, NCI Thesaurus, Pathogen Transmission Ontology and the Infectious Disease Ontology. It uses the Basic Formal Ontology as an upper ontology, but we expect to integrate it into the mid-level Medical Surveillance ontology. See https://code.google.com/p/epidemiology-ontology/.

Activities at FFCUL in the reporting period (year 4 + extension) in Task 3.3:

- Completely redesigned front-end with much improved usability. This is a full redesign. The team hired the services of a professional graphic design to create an Epiwork-specific style. The entire set of interactions on the website have been re-created using the knowledge acquired with the implementation of the previous versions of the EM.
- The final version the p Epidemic Marketplace was presented in the Epiwork workshop in Digital Epidemiology in Torino (may 2013).
- WP3/WP4 collaboration on the integration of the computational platform with the data platform.
- Continued design of new access control system to be implemented in the EM.
- Integration of GleamViz with the Epidemic Marketplace. We jointly developed an interface that will enable GleamViz upload simulation results as datasets to the EM platform to be shared by the community.
- Developed the software for integration with Influenza.net (WP5)

Activities at FFCUL in the reporting period (year 4 + extension) in Task 3.4:

- Continued work on planning and monitoring the operation of the EM. Results are documented in the final WP3 deliverable (D3.6).

2.3 *Effort Allocation*

The effort allocated by LASIGE to WP3 in the reporting period (and previous years) is as follows:

WP3 **FFCUL** **Reporting Period**	**Effort** (p.m) Non-Perm + Perm =Total
Year 1: February 1, 2009 to January 31, 2010	19.38 + 4.53 = 23.91
Year 2: February 1, 2010 to January 31, 2011	32.55 + 6.43 = 38.98
Year 3: February 1, 2011 to January 31, 2012	40.86 + 8.39 = 49.26
Year 4: February 1, 2012 to January 31, 2013	**51.49 + 3.79 = 55.27**
Extension: February 1 to July 31, 2013	**22.41 + 2.09 = 24.49**

The commitment to the project in the first year was 23.91 persons.month, 19.38 p.m from technicians, and 4.53 p.m from permanent staff.

The commitment to the project in the second year was 38.98 persons.month, 32.55 p.m from technicians, and 6.43 p.m from permanent staff.

The commitment to the project in the third year was 49.26 persons.month, 40.86 pm from technicians, and 8.39 pm from permanent staff.

The commitment to the project in the forth year was 55.27 persons.month, 51.49 pm technicians, and 3.79 pm from permanent staff.

The commitment to the project in the extension was 24.49 persons.month, 22.41 pm technicians, and 2.09 pm from permanent staff.

The total effort by FFCUL/LASIGE on WP3 was 191.91 persons.month, 166.69 pm from technicians and 25.22 from permanent staff. This is in line with the effort planned on the revised DoW for WP3, 167 pm.

2. Work in WP4 — Epidemic Modelling Platform

This task is lead by ISI, with a total FFCUL contribution of 19 persons.month (12 hired + (7) academics).

Activities at FFCUL in forth year and extension of the project in WP4:

The LASIGE work on WP4 is related to the integration of the Computational and Data Platforms of Epiwork, which has been accounted in detail in the description of activities of LASIGE on WP3. Below, we mention the most significant of these activities:

- Mário Silva, João Zamite and Carlos Santos participated in the March 2012 project review in Brussels and prepared a joint demo of the GleamViz integration with the Epidemic Marketplace.
- Daniela Paolotti visited LASIGE for one week in May for planning the integration of the GLEaMviz platform and the Epidemic Marketplace and discuss the integration of the EM (WP3) with Infuenzanet (WP5), regarding the organization of datasets for use for computational epidemiology tools.
- Upgrade to the EM implied additional work on the integration of the Computational Platform tools, in particular GleamViz.
- Mário Silva participated in the Epiwork project meeting in May 2013 in Torino and presented the integration work with WP4 at the Epiwork organised International Workshop on "Digital Epidemiology"

2.4 *Effort Allocation*

The effort allocated to WP4 in the in the reporting period (and previous years) is as follows:

WP4 FFCUL Reporting Period	Effort (p.m) Non-perm + Perm =Total
Year 1: February 1, 2009 to January 31, 2010	5.72+0.50 = 6.22
Year 2: February 1, 2010 to January 31, 2011	5.05 + 0.40 = 5.45
1.4.1.1 Year 3: February 1, 2011 to January 31, 2012	1.4.1.2 7. 31 + 2.53 = 9.84
Year 4: February 1, 2012 to January 31, 2013	**3.36 + 2.54 = 5.90**
Extension: February 1 to July 31, 2013	**0.30 + 0.13 = 0.43**

The commitment to the project in the first year was 6.22 persons.month, 5.72 p.m from technicians, and 0.50 p.m from permanent staff.

The commitment to the project in the second year (the reporting period) was 5.45 persons.month, 5.05 pm from technicians, and 0.40 p.m from permanent staff.

The commitment to the project in the third year was 9.84 persons.month, 7.31 pm from technicians, and 2.53 pm from permanent staff.

The commitment to the project in the forth year was 5.90 persons.month, 3.36 pm technicians, and 2.54 pm from permanent staff.

The commitment to the project in the extension was 0.43 persons.month, 0.30 pm technicians, and 0.13 pm from permanent staff.

The total effort by FFCUL/LASIGE on WP4 was 27.83 persons.month, 21.74 pm from technicians and 6.09 from

permanent staff. This is in line with the effort planned on the revised DoW for WP4, 30pm.

The reported effort by technicians reflects the splitting of the effort dedicated to setting-up the hardware and base software of the Epiwork infrastructure in Lisbon between WP3 and WP4 in the first year, and, in the second and third years, the development of the SimpleEMClient for synchronizing local folders with EM streams and work on demonstrating and supporting the integration of GleamViz with the Epidemic Marketplace in cooperation with ISI.

Work in WP7 — Management

This task is lead by ISI.

FFCUL Effort in this task: 4 persons.month.

Activities at FFCUL in the fifth semester of the project in WP7:

1. Data collection activities for project tracking, preparation of the 5th semester report and 3rd year reports, management presentations.
2. Mário Silva and Dulce Domingos attended the March 2011 Project Review in Brussels and the third Epiwork Meeting in Courmayeur, Italy in January 2012.

2.5 *Effort Allocation*

The effort allocated to WP7 in the in the reporting period (and previous years) is as follows:

WP7 **FFCUL** **Reporting Period**	**Effort** (p.m) Non-Perm + Perm = Total
Year 1: February 1, 2009 to January 31, 2010	0.0 + 1.10 = 1.10
Year 2: February 1, 2010 to January 31, 2011	0.0 + 0.96 = 0.96
Year 3: February 1, 2011 to January 31, 2012	**0.0 + 0.99 = 0.99**
Year 4: February 1, 2012 to January 31, 2013	**0.0 + 0.55 = 0.55**
Extension: February 1, 2013 to July 31, 2013	**0.0 + 35 = 0.35**

The effort dedicated to the project was 100% contributed by permanent

staff, as planned.

The total effort by FFCUL/LASIGE on WP7 was 3.95 persons.month. This is in line with the initially estimated effort on the DoW for WP7, 4pm.